An integrative approach to the relations of prosody to discourse: towards a multilinear representation of an interface network.

Albert Di Cristo, Cyril Auran, Roxane Bertrand, Catherine Chanet, Cristel Portes

An integrative approach to the relations of prosody to discourse: towards a multilinear representation of an interface network.

Albert Di Cristo[*], Cyril Auran[**], Roxane Bertrand[***], Catherine Chanet[****], Cristel Portes[*****]

Abstract

The study of the relations between prosody and discourse, as we view it, requires the development of an analysis tool adapted to the description of complex interface networks comprising the heterogeneous data from both prosodic and inference-driven text-centred analyses. Such a tool can be designed as a multi-linear grid, the methodological and theoretical bases of which are exposed in this paper.

1.Introduction

This paper presents the theoretical foundations and the methodology which characterize our approach to the relations of prosody to discourse. These relations are considered to take part in both the production and the interpretation processes inherent in spontaneous spoken discourse in French. Although this approach relates to the cognitive linguistics stream inspired by researchers attached to the « situated cognition » paradigm or to the « pragmatic perspective » (Verschueren, 1994), the formal description of linguistic objects constitutes one of its major centres of interest. More specifically, our approach intends to pay the greatest attention to conceptual and discourse factors in order to account for the layout of linguistic (both syntactic and prosodic) objects and for the cognitive strategies which constitute the basis for the online encoding and treatment of discourse. This theoretical position thus aims at favouring exchanges between the traditionally antagonistic linguistic trends of formalism and functionalism (Newmeyer, 1998).

In this perspective, our basic hypothesis is that the general apparatus allowing to account for the complex and non-bijective nature of the form-function relations can be represented with a meta-grammar called Ecological Grammar (Di Cristo, 2000). It is important to note that the term « grammar » is not used here in its classical sense of « formal grammar » (or « competence grammar »), but in the more general sense of a descriptive system of the existence and functioning of a natural language. The unusual association of the terms « ecology » and « grammar » (taken from Lambrecht), means that we regard « situated discourse » as an ecosystem, a mode of expression which constantly adapts to its setting. The term « ecology » also seems to suit the flexibility and adaptability of prosody in the flow of discourse. More generally, this term makes room for a link with cognitive sciences, one of the major tenets of which is that intellectual capacities and perceptive functions are processing systems for environment data.

Ecological Grammar is a meta-grammar which interfaces a contextualisation grammar with a grammar of linguistic expressions. The contextualisation grammar (or grammar of pragmatic competence in discourse) is based on several principles from diverse theories and models such as Gumperz's contextualisation model, Sperber & Wilson's Relevance Theory, Lambrecht's and Vallduví's conceptions of information structure, and Grize's Schematization Theory.

The grammar of linguistic expressions consists of the elements and the constructions of the syntactic and the prosodic systems, « summoned » and configured by the contextualisation grammar.

The structuring of the two grammars originates in an integrative approach to the relationship of language to communication. It proposes to do away with the closed strategy of reductionism and the thesis of the immanence of linguistic facts, in favour of the open strategy of contextualism, which strives to explain phenomena taking into account both their internal properties and external facts. As an example, this implies that we try to account for the organisation of discourse prosody by explaining how the constraints inherent in the prosodic systems (which constitute evidence for

[*] Laboratoire Parole et Langage, UMR 6057, CNRS, Aix-en-Provence : albert.dicristo@lpl.univ-aix.fr
[**] cauran@wanadoo.fr
[***] roxane.bertrand@lpl.univ-aix.fr
[****] catherine.chanet@lpl.univ-aix.fr
[*****] cristel.portes@lpl.univ-aix.fr

the relative autonomy of these systems) interact with those related to the processes and stakes in discourse. The interface of the contextualisation grammar with the grammar of linguistic expressions also constitutes a complex interface, or rather an interface network, which goes beyond the traditional interfaces (prosody/syntax, prosody/semantics, prosody/pragmatics), and the study of which is related to the scientific paradigm of a generalised optimality theory. Such a perspective relies on a heurism which aims at listing, in an inductive mode, a set of regularities which might in turn shape a model of the relations of prosody to discourse (a model still to be developed) or enrich existing discourse models which do not take prosody into account.

We consider that the interpretation of the interface network which constitutes the core of Ecological Grammar requires multi-linear modes of representation, applied both to prosodic marks and to textual and/or discourse clues. We also argue that the relations between these representations can be adequately established through specific grids. This paper details the tenor of these grids, and exposes the principles which led to their development before providing several illustrations in the last part of this analysis.

2. Development principles of the text-discourse analysis grid

Spoken discourse combines a text and a prosody. The linguistic decoding and the pragmatic inferences which contribute to the interpretation of spoken discourse (the understanding of the messages and of the speakers' intentions) are based upon the interactive processing of these components. Since we a priori ignore the nature of these relations, we consider it to be legitimate to represent these components on separate grids: the text-discourse grid and the prosodic grid.

The development of the text-discourse grid, based solely on data from the text, induces the study not only of its syntax but also of the elements which confirm the functional heterogeneity of discourse (Bertrand, 1999) present in the text. It is thus of major importance to mark the information related to the signalling of logical relations between discourse constituents (such as topics, sub-topics, clauses, etc.) which constitute the very frame of discourse macro-structure. But it is equally important to mark the discourse activities which fit into this frame. In our approach, these discourse activities are apprehended with a widened conception of contextualisation (*text*-, *ego*- and *alter*-centred) which takes into account aspects related to information contextualisation, enunciation, and to the expression of affect and interaction. In short, the text-discourse grid aims at making a multi-linear inventory of the marks from different levels which are present in the text and allow inferences on the discourse itself, seen as the activity generating the text as a discourse product.

These levels mainly concern the lexicon, syntax, enunciative phenomena, referential phenomena and their combinations, as well as data provided by the implementation of diverse discourse models such as RST (Mann & Thompson, 1988) or the Geneva Model (Roulet et *al*., 2001). Several lines in the grid can be devoted to the record not only of the agreements and discrepancies between marks from graphically superior levels but also of their interpretations. These interpretations can concern argumentative as well as structural aspects of discourse.

We lack space in this paper to describe the structure of a text-discourse grid as presented elsewhere (Portes *et al*., 2002, Di Cristo *et al*., 2003). However, it is important to note that this grid is divided into two parts, devoted to the mark-up of phenomena within utterances and between utterances respectively. The first part constitutes a record of lexical, syntactic, semantic and enunciative « micro-information ». The second part mainly focuses on phenomena building up discourse coherence on the basis of referential, argumentative and structural elements. It is precisely in this part of the grid that analyses derived from the discourse models mentioned above may be recorded.

It is also important to insist on the fact that these text-discourse grids are modular and that it is absolutely not necessary that all the lines of the maximal grid should be used for the analysis of a particular example.

3. Development principles of the prosodic grid

The development of the prosodic grid obeys the same general principal of multi-linear representation. This may be justified by several reasons. First, in our approach, prosody, which is often reduced to intonation, is conceived as a complex system consisting of three orders of structure: metrical organisation (which governs the attribution of

prominences and rhythmic phrasing), tonal organisation (which supervises the sequential organisation of tonal segments and melodic patterns) and temporal organisation (which controls syllabic duration, pauses and speech rate).

Secondly, we consider that each order of structure is organised along two dimensions : a syntagmatic dimension and an orthogonal dimension. The syntagmatic dimension of tonal organisation can thus be interpreted in terms of melodic marks and intonation patterns, whereas its orthogonal dimension refers to superpositional phenomena of register, span and downtrend. In the same way, variations of speech rate will be regarded as elements of the orthogonal dimension of temporal organisation, marked in the syntagmatic dimension by local variations of syllabic durations and pauses.

Thirdly, taking into account the bi-dimensional nature of the prosodic orders of structure allows us to interpret them in terms of local, global and iterative properties, which justifies the use of a multi-linear representation of the prosodic marks implementing these properties.

Fourthly, we think it is important to differentiate, in the prosodic grid, the mark-up of formal entities and that of functional ones, such as boundaries, emphasis and contrast (see below).

Lastly, numerous authors have insisted on the fact that one can not establish direct relations between prosodic functions and the physical parameters which support them since one or more intermediate levels of representation are necessary.

The multi-linear coding of prosody presented in the analysis grid, confronted with the multi-linear mark-up of text-discourse phenomena and cues mentioned in the previous section, actualizes the outcome of a process based upon the use of several complementary approaches. The first approach relates to the perceptual annotation of a set of audible phenomena most of which are not, at present, identifiable through automatic processing. This subjective annotation concerns accentuation (2 categories : emphatic and non-emphatic), boundaries (2 categories : conclusive and non-conclusive, with an index from 1 to 4 for each category), registers (3 categories : normal, upgraded, downgraded), span (3 categories : normal, widened, reduced), syllabic duration (3 categories : short, long, very long), speech rate (3 categories : normal, fast, slow), pauses (2 categories : long and very long), breaths and laryngalizations.

The other approaches used are objective in kind and respectively concern normalisation of phonemic duration and of fundamental frequency, automatic detection of pauses, transcoding of raw (physical) data into perceptual data and automatic (or semi-automatic) coding of several prosodic elements. At the present stage of our research, automatic and semi-automatic codings only apply to tonal organisation through modelling of the fundamental frequency curve and marking of intonation using the INTSINT intonation alphabet (Hirst & Di Cristo, 1998). These processes consist in deriving a surface phonological representation of the tonal organisation from the acoustic signal; this representation allows the identification of tonal marks and intonation patterns, the anchoring points which play an active part in the laying out of discourse.

The approaches used to fill in the multi-linear representation of prosody can obviously be regarded as heterogeneous, in the sense that they concern both objective procedures and subjective ones. As far as the latter are concerned, it is equally blatant that they jointly involve the marking of formal phenomena (such as syllabic duration or registers, for example) as well as functional ones (such as boundaries and accent categories, for example). This heterogeneity, however constitutes but a transitory stage, current research aiming at developing automatic annotation of syllabic duration, and speech rate, register and span variations. Moreover, as far as we know, no efficient procedure for the automatic detection of accent categories and boundaries levels has been developed. This seems to justify the relevance of the subjective annotation of boundaries; indeed, one can not make any a priori judgement on what part is played in the identification of these boundaries by prosodic cues only or by a set of elements jointly or alternatively provided by prosody, syntax and pragmatically motivated inferential processing. Through the use of the complete grid, the mapping of prosodic and text-discourse data should shed some light on this issue which approaches dedicated to the description of a single interface (such as the intono-syntactic approach for example) can not satisfactorily solve.

A sample prosodic grid begins with a speech wave representation of the signal (**Osc**). All the following lines are consequently aligned with this representation which constitutes the first line in the entire grid.

The second line (**Momel**) represents the frequency curve modelled as a smooth continuous curve joining sequences of target-points. This modelled curve is interpreted as a « phonetic representation » of tonal organisation (Hirst *et al.*, 2000).

These first two lines also contain information about the segmentation into tone units (UI), numbered according to their order of appearance. Line **Ortho** displays the orthographic transcription of the text aligned with the signal and deprived of any punctuation mark.

The following four lines are devoted to the categorial coding of tonal organisation. the **INTSINT** line consists in the categorial coding of target-points of the phonetic representation (this coding being interpreted as a surface phonological representation of tonal organisation). Lines **Dd** (« *downtrends* »), **ReSp** (« *Register Span* ») and **ReLe** (« *Register Level* ») correspond to the coding of elements in the orthogonal dimension of tonal organisation (downtrends effects, span variations, and register level).

Line **Brea** (« *Breathing* »), which records breaths, comes before the three lines devoted to the coding of phenomena related to temporal organisation: pauses : **Pau**, syllabic duration : **Length** (N : « *Normal* », R : « *Reduced* », Le : « *Lengthened* » et Xle : » *Extra-lengthened* »), and speech rate variations : **Tempo** (N : « *Normal* », Fa : « *Faster* » et Slo, : « *Slower* »).

The following two lines, devoted to the coding of metrical organisation, respectively concern prominences : **Proe** (aligned with the text) and their metrical strength levels : **Met**, from 1 to 3.

The last lines in the prosodic grid are devoted to functional marks: Accent, Boundary, Emphasis and Contrast. As we can not take for certain that these functions can all be exclusively attributed to prosody, it seems legitimate to place their coding in a watershed position in the grid, between prosodic annotations and text-discourse annotations.

4. Conclusion

The study of the relations between prosody and discourse can be considered from the viewpoint of the description and modelling of a complex interface network allowing the establisment of non-bijective relations between heterogeneous forms and functions; as such, it requires the development and implementation of an adpated and evolutionary tool. We consider that the prosodic and text-discourse grids the principles of which have been presented in this paper fulfil these conditions. Theses grids allow a mosaic approach which is characteristic of developing paradigms. This means that they must primarily be regarded as the illustration of a heurism aiming at developing a knowledge base about the role of prosody in the elaboration and processing of discourse.

5. Illustrations

The vertical analysis of the relations between the different text-discourse and prosodic primary elements favours the description of some interesting interface effects:

UI1 ends with a specific final lowering pattern. Such a pattern could signal the end of a paratone, but it has also been described as a prosodic mark typical of headlines (Wichmann 2000 mentions a « citation countour »). This latter interpretation is reinforced by a number of cues: on the informational and semantic levels, indeed, UI1 introduces the main topic to be developed in the rest of the extract; on the prosodic level, a higher register level, together with an expanded span are characteristic elements of a beginning paratone.

UI2 et UI3 display comparable final continuative patterns (equal heights for T1 and T2, line 6). This close copy of patterns prosodically materializes the list effect announced in the main topic (line 22).

The declination effect signalled by the progressive lowering of the T targets (T2>T3>T4, line 6) corresponds to the non-defining relative clause (line 18) thus identified as a homogenous entity dealing with the sub-topic (line 21) co-extensive with UI3.

The echo effect between the two accentual arcs (half-circles on the figure) underlines the coherence in the information concerning the sub-topic.

Although the last intonation unit (UI6) fully takes part in the declination effect which homogenizes the non-defining relative clause, a notable rise in accentual density compared with the preceding UIs, accompanied by a final lowering configuration, foretells the impending termination of the turn.

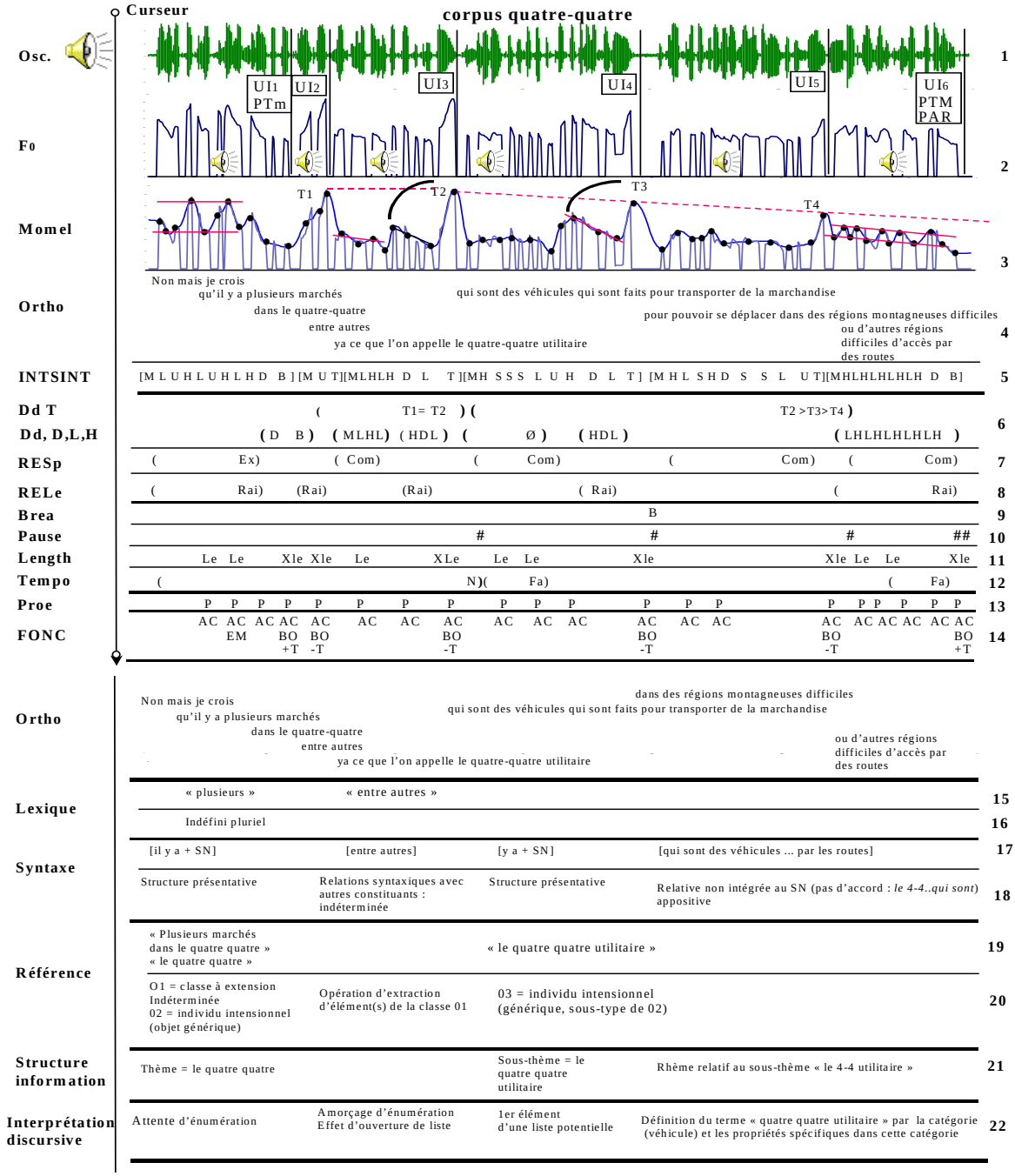

Figure 1: example of a multi-linear grid

References

Bertrand R, 1999, *De l'hétérogénéité de la parole : analyse énonciative de phénomènes prosodiques et kinésiques dans l'interaction interindividuelle*, thèse de doctorat, Université de Provence.

Di Cristo A., 2000, « Une Grammaire Ecologique comme cadre interprétatif de la prosodie de la parole », 3rd All-Nordic and Baltic Congress on Semiotics (June 2000), Imatra, Finland.

Di Cristo A., Auran C., Bertrand R., Chanet C., Portes C. & Régnier A., à paraître, « Outils prosodiques et analyse du discours », CILL 28, Université de Louvain.

Grize J.B., 1996, *Logique naturelle et communication*, Paris, Presses Universitaires de France.

Gumperz J.J., 1992, Contextualization and understanding, in *Rethinking Context: Language as an Interactive Phenomenon*, A. Duranti & C. Goodwin (eds), Cambridge, Cambridge University Press, p. 229-252

Gumperz J.J., 1997, Communicative competence, in *Sociolinguistics : A Reader and Coursebook*, N. Coupland & A. Jaworski (eds), Houndmills, N. Hampshire, Macmillan, p. 39-48.

Hirst D.J. & Di Cristo A., 1998, *Intonation Systems*, Cambridge (UK), Cambridge University Press.

Hirst D.J., Di Cristo A. & Espesser R., 2000, « Levels of representation and levels of analysis for the description of intonation systems », in *Prosody : Theory and Experiment*, M. Horne (ed.), Dordrech, Kluwer, p. 51-88.

Lambrecht K., 1994, *Information Structure and Sentence Form: Topic, Focus and the Mental Representation of Discourse Referents*, Cambridge, Cambridge University Press.

Lave J., 1988, *Cognition in practice*, Cambridge (UK), Cambridge University Press.

Mann W.C. & Thompson S.A., 1988, « Rhetorical structure theory: toward a functional theory of text organization, Text, 8 (3), p. 243-281.

Newmeyer F.J., 1998, *Language Form and Language Function*, Cambridge (MA), MIT Press.

Portes C., Rami E., Auran C. & Di Cristo A., 2002, « Prosody and discourse : a multi-linear analysis », proceedings of the 1st International Conference of Prosody (Aix-en-provence), p. 579-582.

Roulet E., Filliettaz L. & Grobet A., 2001, *Un modèle et un instrument de l'analyse du discours*, Bern, Peter Lang.

Sperber D. & Wilson D., 1986, *Relevance: Communication and Cognition*, Cambridge, Harvard University Press.

Verschueren J., 1994, « The pragmatic perspective », in *Handbook of Pragmatics*, J. Verschueren, J. Östman & J. Blommaert (eds), Amsterdam/Philadelphia, John Benjamins, p. 1-19.

Vallduví E. 1991, *The Informational Component*, New York, Garland.